DIGITALLY-GENERATED LIGHTING
FOR VIDEO CONFERENCING APPLICATIONS

CROSS-REFERENCE TO RELATED APPLICATIONS

[01] This non-provisional application claims the benefit of priority to U.S. Provisional Application Serial No. 60/265,842 filed February 1, 2001, which is expressly incorporated by reference.

FIELD OF THE INVENTION

[02] The present invention generally relates to image processing, in particular to a system and method of digitally generating light for electronic video conferencing over telecommunication networks.

BACKGROUND OF THE INVENTION

[03] In the last few years, the development of high quality multimedia and the availability of powerful computing platforms ability to handle video and audio in real-time has increased the interest in video and audio applications. Although, video coding and delivery schemes are available to the users, the quality of the picture is generally undesirable due to bad lighting conditions caused in part by the poor quality of available lighting apparatus. Conventional video conferencing apparatus have certain drawbacks regarding the transmitted picture quality. In particular, video conferencing apparatus very often transmits underexposed participants, which degrades the viewing experience and the quality of the conference. User-provided camera controls are generally insufficient for compensation of bad lighting conditions. In addition, it is undesirable to have numerous special lights at the physical conference location. These special lights may be uncomfortable to the participants due to additional heat given off by the lights. Other drawbacks include, the lack of sufficient electrical power to work the lights, and the inability to control the light parameters. Even if one was to try spot lighting

the scene with physical lights, the trial-and-error process is highly inefficient and costly.

[04] An approach to re-lighting has been developed as an extension of computer augmented reality (CAR). In such methods employing CAR, a set of static photographs taken in controlled light conditions or a video sequence are reconstructed three-dimensional geometrically. Several vision techniques are used for the scene reconstruction, such as camera calibration, mosaicing, computation of the epipolar geometry, which results in a polygonal reconstruction of the scene. Light exchanges among objects in the scene are computed and illumination textures coming from real and synthetic lights are modeled and reintroduced in the scene. While these CAR systems provide realistic effects of re-lighting, these CAR systems have certain drawbacks and undesirable features. These CAR systems are complex, non-real time based, and require entire geometrical reconstruction of the entire scheme being re-lighted. These CAR systems do not accommodate video conferencing applications, nor real-time applications. Also, the CAR systems do not dynamically adjust a light source illuminating an object in the scene when the object is moving in real-time. Thus, what is needed is a system and method for improving image and talking head sequences.

SUMMARY OF THE INVENTION

[05] The present invention pertains to a system and a method to improve the lighting conditions of a real scene or a video sequence. In one aspect, digitally generated light is added to a scene for video conferencing over telecommunication networks. A virtual illumination equation is implemented that takes into account light attenuation, lambertian and specular reflection. In another aspect, the present invention modifies real light sources intensities and inserts virtual lights into a real scene viewed from a fixed viewpoint.

[06] In a first aspect of the present invention, there is provided a computer-implemented method of the digitally illuminating an object in real-time. The

-2-

method features capturing an image of an object, providing at least a virtual light source for illuminating the object within the image, extracting a surface position of the object, illuminating the object at least at the extracted surface position with the virtual light source, and displaying an illuminated object within the image. In further aspects of the present invention, a two-dimensional plane of the object is created in which the two-dimensional plane illuminated with the virtual light source. In another aspect, during illuminating of the object, a diffused light component and a specular lighting component are combined.

[07]   In a second aspect of the present invention, there is provided a method of selectively illuminating a head of a user for an image processing system in real-time. The method features capturing an image of the head of the user, determining the position of the head of the user to obtain position information; generating a model of the head of the user using the position information, applying a synthetic light to a position on the model to form an illuminated model, and combining the illuminated model and the image. In further aspects, the position of the head of the user is dynamically tracked so that an ellipsoid is generated which is representative of the head of the user. In this manner, at least video conferencing applications where typical lighting conditions (i.e. the average office or home environment) are poor can be greatly improved.

BRIEF DESCRIPTION OF THE DRAWINGS

[08]   For a better understanding of the present invention, reference is made to the following description of exemplary embodiments thereof, considered in conjunction with the accompanying drawings, in which like reference numbers refer to like elements, and wherein:

[09]   FIG. 1 is a schematic diagram of a first embodiment of an image processing system according to the teaching of the present invention;

[10]   FIG. 2 is a schematic diagram of a second embodiment of an image processing system according to the teaching of the present invention;

[11]   FIG. 3 is a schematic representation of a lighting model;

[12]   FIG. 4 is a schematic representation of a point light model;

[13]   FIGS. 5A-5D are example three-dimensional diagrams of a museum lighting model;

[14]   FIGS. 6A-6F are example pictorial representations of illuminated images with a museum model;

[15]   FIG. 7 is a schematic diagram of a method for facial feature tracking analysis;

[16]   FIGS. 8A and 8B are schematic diagrams illustrating head and facial feature tracking on a subject;

[17]   FIG. 9 is a schematic diagram of three-dimensional head tracking model;

[18]   FIG. 10A-10C are example pictorial representations of illuminated images with a museum model and three-dimensional head model;

[19]   FIG. 11 is a generalized flow diagram of a method of digitally illuminating an object; and

[20]   FIG. 12 is a schematic diagram of an example of a video conferencing system for the method of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[21]   To assist the reader in the understanding of the invention and ease of explanation, the specification has been divided into the following sections:

Illumination Models, Still Image Processing Application, and Video Conferencing Application.

[22] An overview of a first embodiment of a video processing system 2 is illustrated schematically in FIG. 1. The first embodiment is used in a pixel based approach. In this approach the system 2 operates directly on the pixels of an input image that is processed. Nevertheless, one of ordinary skill in the art should recognize that at least the pixel color values in the area of interest can be varied to create the lighting effect. The pixel color values include red, green, and blue values, which are conventional components of a RGB signal. Other signals may be used to vary brightness and contrast of the image. A second embodiment of a video processing system 2' is illustrated in FIG. 2. The second embodiment is used with a conventional 3D texture mapper 16'. Texture mapped lighting is applied to the original input image or video. This can improve the visual quality of the illuminated image. Texture mapping is a technique commonly used in electronic video games in a simulated environment.

[23] Nevertheless, both embodiments of the video processing system 2, 2' operate to improve image quality. In general, video processing system 2 and 2' operate by employing a virtual illumination equation that takes into account light attenuation, lambertian and specular reflection. This equation is used in an example embodiment to model a 3D space in which an image or video frame lies on a 2D plane. Virtual lights are placed in the 3D space and illuminate the 2D image. In an alternative embodiment for talking head video sequences, the head of a subject is modeled as a three-dimensional object, such as an ellipsoid. Then virtual lights are used to illuminate the head of a person in front of a camera (shown in FIG. 12).

[24] Video processing system 2 includes several software or program procedural components that execute programmed instructions for specific predetermined purposes. The program procedural components includes some or all of the

following modules – an image capture module 4, a static lighting module 6, a mixing module 8, a tracking module 10, a modeling module 12, a dynamic lighting module 14, and a video rendering module 16. Referring to FIG. 2, video processing system 2' also includes image capture module 4, static lighting module 6, tracking module 10, modeling module 12, dynamic lighting module 14, and texture mapper 16' includes a rendering portion. Each module will be described in detail herein.

[25]   A brief overview of the function of each module is described below. Image capture module 4 receives video scene data from a camera or any suitable digital video camera for processing on the system. Static lighting module 6 receives the video scene data and inserts synthetic or virtual lighting into the video scene data. Mixing module 8 receives data from the static lighting module and dynamic lighting module 14 so as to digitally combine virtual lighting and real lighting for realistic lighting enhancement.

[26]   Tracking module 10 provides position information when the heads of persons seated in front of a video conferencing camera moves. Modeling module 12 creates a three-dimensional model of the head of person seated in front of the camera so as to form an ellipsoid model. Dynamic lighting module 14 receives position information and moves the lighting to correspond to the movement of the head being tracked. Video rendering module 14 displays the processed video data from the mixer module 8. It should be understood that the processed video data can be also transmitted to a destination video conferencing device (see FIG. 12).

[27]   Shown schematically in FIG. 1, video processing system 2, 2' may operate in a computing environment of a conventional general-purpose computer, including but not limited to a mini-computer, a high-speed computer workstation, a personal computer, or a laptop computer. Hardware components of system 2 may include a central processing unit (not shown), a system memory (not shown), and a system bus (not shown) that couples

various computer system components. The central processing unit may be one or more suitable general-purpose microprocessors used in a conventional computer. The system may use any number of conventional operating systems, such as MICROSOFT WINDOWS®, WINDOWS NT®, WINDOWS XP®, UNIX®, or LINUX® or other variations.

[28] Referring to FIG. 1, mixing module 8 receives data from the static lighting module 6 and dynamic lighting module 14 so as to digitally combine virtual lighting and real lighting for realistic lighting enhancement. The mixing module employs standard conventional numerical smoothing algorithms for clipping and color saturation control as known in image processing. As one can appreciated any number of methods can be implemented for mixing the virtual lighting and real lighting for display.

[29] Further into the details of the inventive image processing system, static lighting module 6 and dynamic lighting module 14 implement a virtual illumination algorithm which may take into account light attenuation, lambertian and specular reflection for enhancing the image as described in the Illumination model section below.

ILLUMINATION MODEL

[30] The most relevant factors that account for change in appearance on an object due to a variation in lighting conditions is reflection and shadow. Reflection can be classified having a diffuse component, which is typical of non-shiny surfaces, and a specular component, which is typical of shiny surfaces. Objects illuminated solely by diffusely reflected light exhibit an equal light intensity from all viewing directions. While specular reflection instead, depends on the viewing direction and is at a maximum value along the reflection direction. A bright highlight called a specular highlight appears from viewing directions near to the reflection direction.

[31] Shadows can be classified as a self-shadow, which is dependent on the surface normal vector and the lighting direction. Self-shadow appears where an object surface does not see the lighting direction. Otherwise, a shadow can be a cast-shadow that depends from the overall illumination and it is observed where other objects occlude the light. So for a given point on an image a simplifed illumination model can be represented as $I = I_a + I_d + I_s$ for regions of non-occluded light and $I = I_a$ for regions with occluded light, where I_a accounts for the intensity of a diffused, non-directional light source called ambient light, I_d accounts for diffused reflection contribution of the surface at that given point and I_s refers to the specular reflection component. Such a light model has been extensively studied and it is commonly known as the Phong model. In addition, the light model can also be expressed by $I = I_a + I_d$ and is known as Lambertian reflection model. It is possible to show that for a given point on a surface, the illumination provided by i distinct lights is given by the following general illumination equation:

$$I\lambda = I_a k_a O_d\lambda + \Sigma^n_i \, I_p\lambda_i[k_d O_d\lambda(N \bullet L_i)) + K_s O_s\lambda(R_i \bullet V)^{m_i}] \qquad (1)$$

[32] The terms in the general illumination equation are defined as follows: Ia is the intensity of ambient light, ka is the ambient-reflection coefficient, $Od\lambda$ and $Os\lambda$ are the diffuse and specular color of the surface of an object being illuminated. The term N is the surface normal vector at the point of illumination and Li is a normalized vector indicating the direction to the ith light source (see FIG. 3). The term $I_{p\lambda_i}$ is the intensity of the ith light source, and kd is the diffuse coefficient or the level of diffuse reflection of that point on the surface. The exponent m_i is called the specular reflection exponent. It determines the speed at which the specular reflection diminishes as the viewing direction moves away from the direction of reflection. In other words, it defines the size of the specular highlight. When the value of m_i is small, the size of the specular highlight is large; as m_i increases, the size of the specular highlight shrinks. Moreover, values of m_i between 0 and 1 reduce the amount of

-8-

specular reflection. The specular reflection coefficient ks controls the overall brightness of the specular highlight, independent of the brightness of the light sources and the direction of viewing. The vector R is the direction of reflection and V is the direction of viewing as shown in FIG. 3.

[33]   Equation 1 does not take into account light attenuation. In practice, a formula that works well is shown below in Equation 2:

$$A_i = min(\frac{1}{c_1+c_2d_s+c_3d_s^2}, 1)$$
(2)

[34]   The terms of equation 2 are defined as follows: Ai represents the attenuation factor for the source light, $c1$ is a user defined constant that keeps the denominator from becoming too small when the light is close, $c2$ and $c3$ are specific to a light source model. The term ds represents the euclidean distance of the light source from a surface point. Therefore, in summary, in the case of n light source the virtual illumination equation implemented in an embodiment of the present invention can be expressed or defined as:

$$I_\lambda = I_ak_aO_{d\lambda} + \Sigma^n_i A_i I_{p\lambda_i}[k_dO_{d\lambda}(N \bullet L_i) + K_sO_{s\lambda}(R_i \bullet V)^{m_i})]$$
(3)

[35]   It should be recognized equation 3 may be used for a single light source as well ($n=1$). Nevertheless, light sources can be generally classified as a) non directional ambient light that contributes with constant intensity equally to every point in a scene; b) directional light that can be considered as a light source with a given direction that is not subject to any attenuation. In other words the source is infinitely far away from the surfaces it is illuminating; and c) point light sources.

[36]   An ideal point light source emits light uniformly and radially. The illumination that a point light source contributes to any point on a given surface depends on several factors, including source intensity and color, surface orientation and the distance from the point light source. A simple

model for a more realistic non-equally radiating point source is commonly known as a Warn model. The model represents some of the directionality of the light typically used in photographic studios. Referring to FIG. 4, in the Warn model, a point light source with direction L is modeled by a point on a hypothetical specular reflecting surface. The light intensity at a given point on the surface depends from angle γ between L and L'. Assuming that the reflecting surface has a specular coefficient of 1 then the light intensity at a point on the object is given by the following equation:

$$I_p\lambda_i = I_{L'} \cos^p(\gamma) \qquad (4)$$

where I_L is the intensity of the hypothetical light source;

p is the reflector's specular exponent; and

γ is the angle between L and L'.

[37] It should be recognized that the larger the value of p, the more the light is concentrated along L'. This model is used to establish virtual light for different types of spot lights.

STILL IMAGE PROCESSING APPLICATION

[38] Referring to FIGS. 5A-5D, a first exemplary model consists of the scene placed on a 2D plane and virtual lights that are placed in the 3D space and illuminate the scene. Note that a given light can be placed on the image (i.e. z=0). In the exemplary model for static lighting, by way of example, looks loosely similar to the situation in which a photograph hanging on a wall is illuminated with spotlights and diffusers as it happens in museums and art exhibitions. For ease of explanation, this model will be referred to as the "museum model".

[39] The museum model includes several advantages for implementation over conventional image processing: First, the model is computationally efficient

-10-

which reduces computational overhead by using image properties such as brightness and contrast and not requiring a detailed scene analysis as in conventional CAR system. Second, a large variety of lighting scenarios and image effects can be generated with the museum model. Purely by way of example without limitation of the present invention, Figure 5A shows a 3D space (a cube of 19x19x19 units) with a point light source placed in (12,12,10) represented by the dark sphere. Figures 5B, 5C, and 5D illustrate cases of light patterns that the model can generate. Regarding the virtual lighting equation (3), the attenuation parameters of are $c1 = 0.3$, $c2 = 0.3$, $c3=0.3$, and $Od\lambda = Os\lambda=1$. Of course, the specific parameters can be adjusted accordingly by one of ordinary skill in the art.

[40] Advantageously, by placing directional light sources, such as direct and indirect point lights, is it possible to generate complex illumination effects on the 2D plane on which the image lies and selectively improve the perceived quality of the image or video sequence of images. It should be appreciated that the term Ka of Equation 3 represents the contribution to the image brightness. It is possible to estimate Ka from the evaluation of the input picture brightness. Ka can compensate if the input picture has a low brightness value. For every point light placed in the scene the corresponding Ks controls the brightness of the specular highlight. Together with the associated specular coefficient and the light direction can be used to highlight the region of interest of a video (for example, as shown in FIGS. 6B-6F below). The term Kd accounts for the diffuse reflection of the surface and has a more general use.

[41] Purely by way of example without limitation of the invention, image processing results based on the museum model and an original image are illustrated in FIGS. 6A-6F. Accordingly, FIG. 6A represents an original image unprocessed image. Note that the image the face of the subject is generally underexposed. FIG. 6B illustrates a single point light applied to the image with parameters, such as $ka =0.1$, $kd =0.1$, $ks = 0.2$, $n=20$. FIG. 6C shows the results of using two point lights with parameters, such as $ka =0.2$, $kd1$

-11-

=kd2=0.1 ks 1=ks2= 0.2 n1=n2 =2. FIG. 6D illustrates the results after there point lights are located at the top of the head and two sides. In addition, a point light is included that instead of being summed to the image brightness is subtracted. The subtraction from the image brightness is a way to generate shadows. FIG. 6F also has the same three point lights, as in FIG. 6D. Finally for comparison purposes, FIG. 6E represents the original image after processing of brightness and contrast but no addition of virtual lighting as in FIGS. 6B-6D.

VIDEO CONFERENCING APPLICATION

[42]   The museum model embodiment can be extended for talking head sequences, such as video conferencing applications. In general, the present invention applies algorithms that enable real-time head detection and tracking. The output of the algorithms provides the head position in several forms, including a rectangle or an ellipse. Preferably, the ellipse information is used in the present invention to generate a three-dimensional ellipsoid as shown in Figure 9. This three-dimensional ellipsoid is used to apply virtual lighting to the talking head. The present invention is not limited to an ellipsoid, but other complex shapes of the head can be obtained and used to generate refined lighting effects.

[43]   Referring to FIGS. 1 and 2, tracking module 10 provides position information when the heads of persons seated in front of a video conferencing camera moves. The tracking module 10 may be implement in a number of algorithms. In general, the outline of the head is detected to provide location for applying the virtual light. One approach to finding the outline of the head is disclosed in U.S. Patent No. 5,987,154 to Gibbon et al., which is herein incorporated by reference. In the approach, in a series of video images, the head of a user is identified by digitizing sequential images to determine moving objects, calculating boundary curvature extremes of regions in the subtracted image, and comparing the extremes with a stored model of a human head. Therefore, the tracking module 10 is enabled to determine the location of the head in a

dynamic manner. In the approach, the outline of the head is generally an ellipse shape on the image plane of interest.

[44]   Once the outline of the head is determined, the facial features of the head are preferably tracked to refine a three-dimension model for applying the virtual lighting. FIGS. 8A and 8B schematically show the outline of the head of a user with various facial features for tracking. The facial features generally include the subject's mouth, eyes, eyebrows, or nostrils. There are a number of approaches to finding the facial features in the outline of the head. FIG. 7 illustrates a schematic diagram of the face analysis algorithm used to identify the locations of the facial features of a person in front of a video camera. The face analysis algorithm is disclosed in US. Patent No. 6,118,887 to Cosatto et al., which is herein fully incorporated by reference. In sum, the facial features are recognized by a multi-modal approach, combining the result of color segmentation and shape analysis. If desired, the outline of the head could be determined in the color and shape analysis approach as well. Nevertheless, two or more facial features are used to obtain the head pose of the subject in front of the camera. In this approach, if the user is looking to the side, the lighting is different than when the user is looks straight into the camera. This approach provides information of whether the head is titled forward or backward, or even looking side-to-side. The head pose preferably provides the system with additional position information for applying the virtual lighting.

[45]   Referring to FIG. 1 and 2, modeling module 12 preferably creates a three-dimensional model of the head of a person or a portion of a head of the person seated in front of the camera (not shown). Preferably, the three-dimensional model is in the form of an ellipsoid model. It is to be understood that the present invention is not limited to a three-dimensional ellipsoid. It is contemplated that more complex shapes can be modeled for the head of the user. This complex shapes can form a refined surface model. For example, a three-dimensional model of the head can be constructed to closely match the

-13-

surfaces and features of the face. This refined surface model can be advantageous for improved shadow effects and point illumination of the head.

[46]   Nevertheless, position data or position information from the tracking module 10 is passed into the modeling module 12. Preferably, the ellipse information is used to generate a three-dimensional ellipsoid for creating a 3D head model as shown in FIG. 9. As shown in FIG. 9, point lights (1-3) are directed to the ellipsoid and reflections are created. Of course, the number of point lights can be adjusted as desired. For positioning the ellipsoid properly the head pose is obtained from the facial features, at least approximately. The facial features are marked on the ellipsoid and given the position of the features, the ellipsoid is then oriented so that the marked features match the ones observed in the image. The specific independent facial features provides the algorithm with at least two to three features to determine the appropriate orientation of the ellipsoid in which to apply the virtual lighting.

[47]   With further reference to FIGS. 1 and 2, dynamic lighting module 14 receives position information and moves the lighting to correspond to the movement of the head being tracked. In this module, the surface information of the ellipsoid is extracted so that the virtual lighting equation can be applied. The point light sources are in predetermined positions with respect to the ellipsoid and the lighting is mapped accordingly. If desired, predetermined surface properties can be added to the ellipsoid to create reflections, such as a shiny surface, or non-shiny. This can provide for a more realistic illuminated image to improve the perceived quality of the video sequence.

[48]   Purely by way of example without limitation of the invention, the results of the processing of the talking head sequence are shown in FIGS. 10A-10C. FIG. 10A is an original image unprocessed without applied virtual lighting. In FIG. 10B, a clear improvement appears from where the illumination model has been applied without 3D information. FIG. 10C shows even greater improvement

where the ellipsoidal model has been used. The lighting in FIG. 10C appears more natural on the face of the subject in front of the camera.

[49]   FIG. 11 illustrates a flow diagram of a method of digitally illuminating an object, such as a head of a person. In step 100, a camera captures the image. In step 102, the head motion of the subject in front of the camera is tracked in tracking module 10. This process produces position information of the head and facial features. In step 104, a three-dimensional ellipsoid is generated from the position information of the tracking module 10. In step 106, virtual lighting is applied to the ellipsoid and mixed together. In step 108, the image is displayed on the screen or transmitted to another device for viewing the image.

[50]   FIG. 12 is a schematic diagram of a video conferencing system 200 for implementing the video processing system of the present invention. Video conferencing system 200 may include several electronic and software components including a user interface 202, a control circuitry 204, a memory 206, a network interface 208, and video processing system 2 or 2'. Each component will be described in detail herein.

[51]   Referring to FIG. 12, user interface 202 is a component of video conferencing system 200, which sends out-bound video signals and receives in-bound video signals relating to the user of the system. User interface 202 is operatively connected to the control circuitry 204 and includes at least a video camera 210, and a display device 212. Display device 212 provides visual signals to the user in the form of alphanumeric characters, colors, graphics, and video movements. Display device 212 may be a known display device, such as a liquid crystal display, or computer monitor. The display surface of the device 212 may include one or more video windows for viewing the other party to the video conference. Video camera 210 may be any number of suitable video cameras for digital use, such as commercially available web-cameras. If desired, a keyboard 214 may be included with user interface 202. With

continued reference to FIG. 12, control circuitry 204 may include hardware and software that enables to operate across a network (not shown). The control circuitry 204 may include a microprocessor for use on digital networks. In one arrangement, the video conferencing system may operate with commercially available operating software. Control circuitry 204 also includes an operable connection to memory 206. Memory 206 stores computer readable data installed or programmed by a user, including instructions for digitally illuminating an object. Memory 206 can be any type, which provides nonvolatile storage that can be electrically erased and/or reprogrammed.

[52]   Some suitable video conferencing systems can include a personal computer configured with a digital video camera or a videophone. It should be recognized that the video conferencing system may be configured with a standardized family of telecommunication protocols, such as H.323 standard used for real-time multimedia communications on packet-based networks having an Internet Protocol (IP). Nevertheless, other appropriate protocols may be used to facilitate connectivity between the video conferencing systems.

[53]   While the present invention has been described with reference to preferred and exemplary embodiments, it will be understood by those of ordinary skill in the art that various changes may be made and equivalents may be substituted for elements thereof without departing from the scope of the invention. In addition, many modifications may be made to adapt a particular situation to the teachings of the invention without departing from the scope thereof. Different hardware may be used than shown and suggested that may comprise hardware, firmware, or software implementations of the present invention. Therefore, it is intended that the invention not be limited to the particular embodiments disclosed, but that the invention include all embodiments falling within the scope of the appended claims.